

Workshop interne BI4people  
20/06/2023

**BI4People WP3 :**

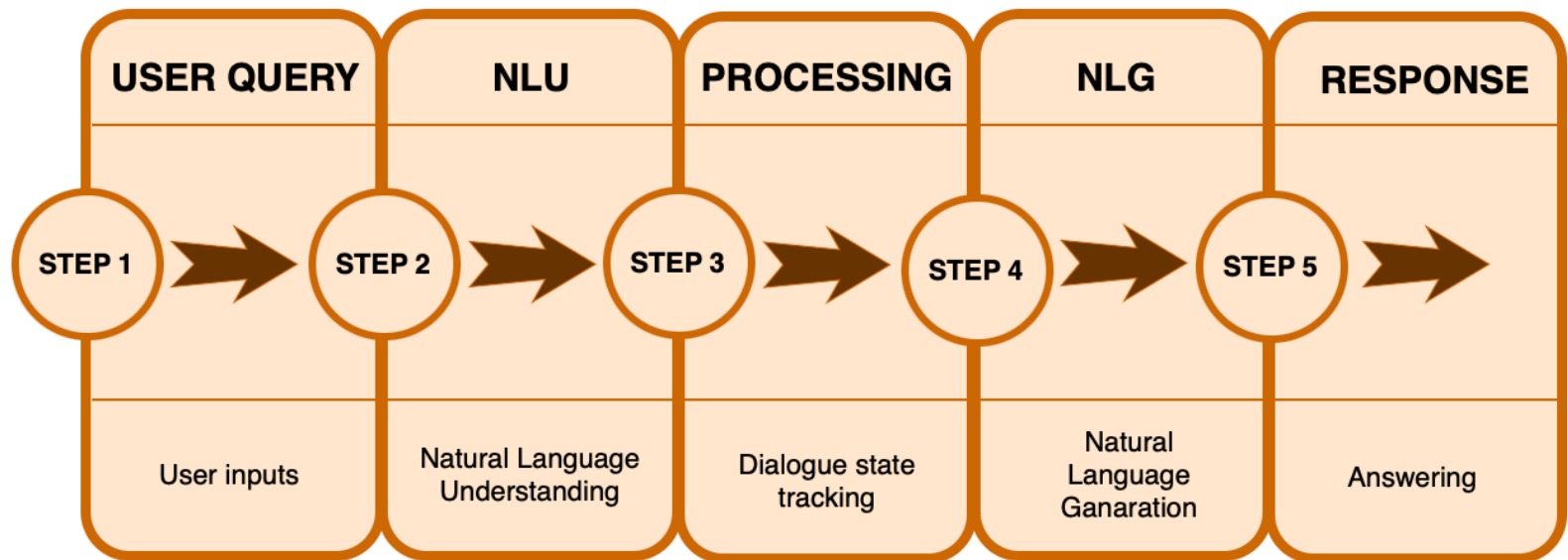
**Collaborative BI Virtual Assistant**

Olga Cherednichenko  
ERIC, Univ. Lyon2

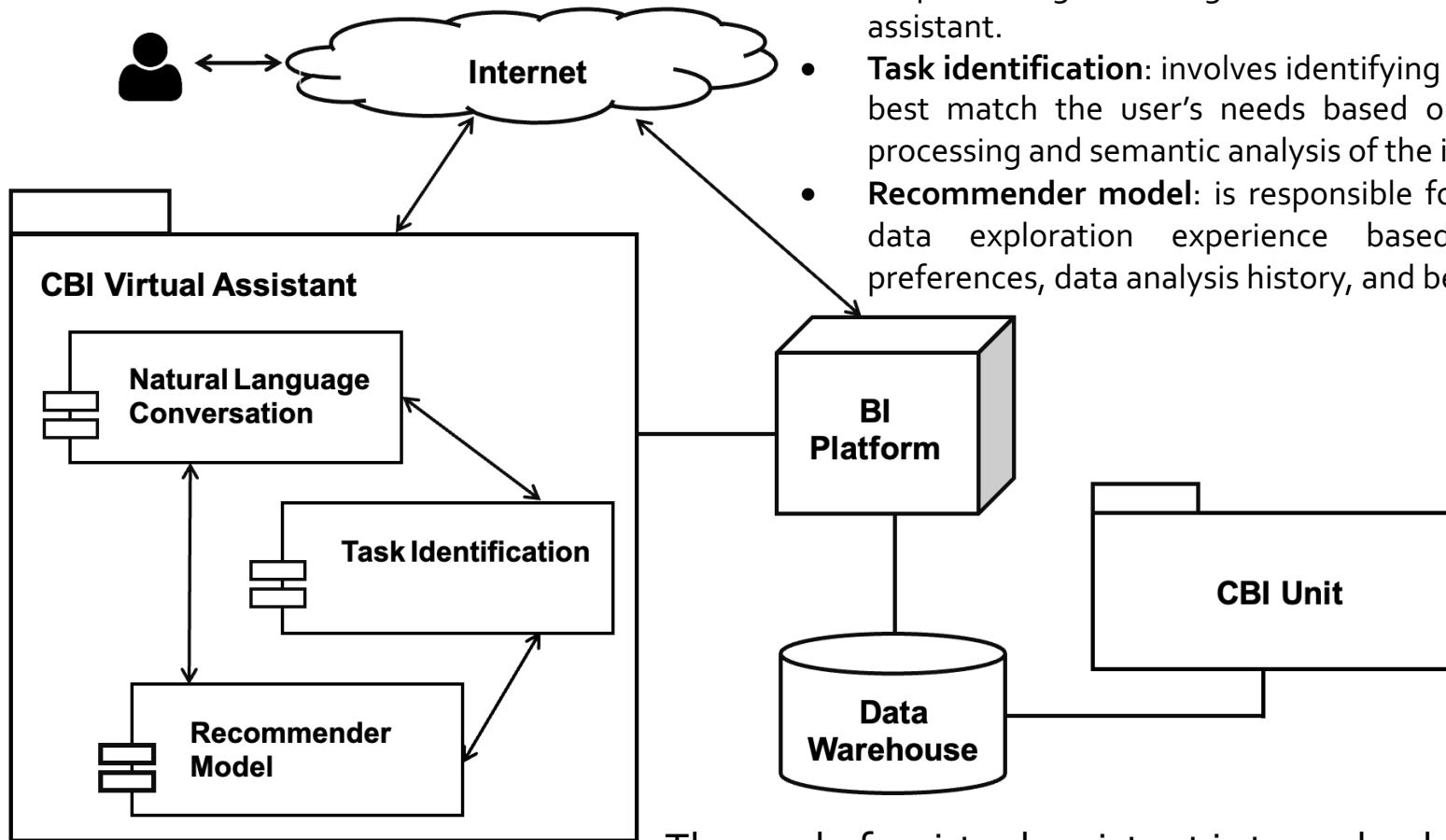


# The general pipeline for conversational agent

**Conversational agents** (CAs) are computer programs designed to engage in natural conversations with human users. They can be categorized as either chatbots for informal chatting or task-oriented agents for providing users with specific information related to a task



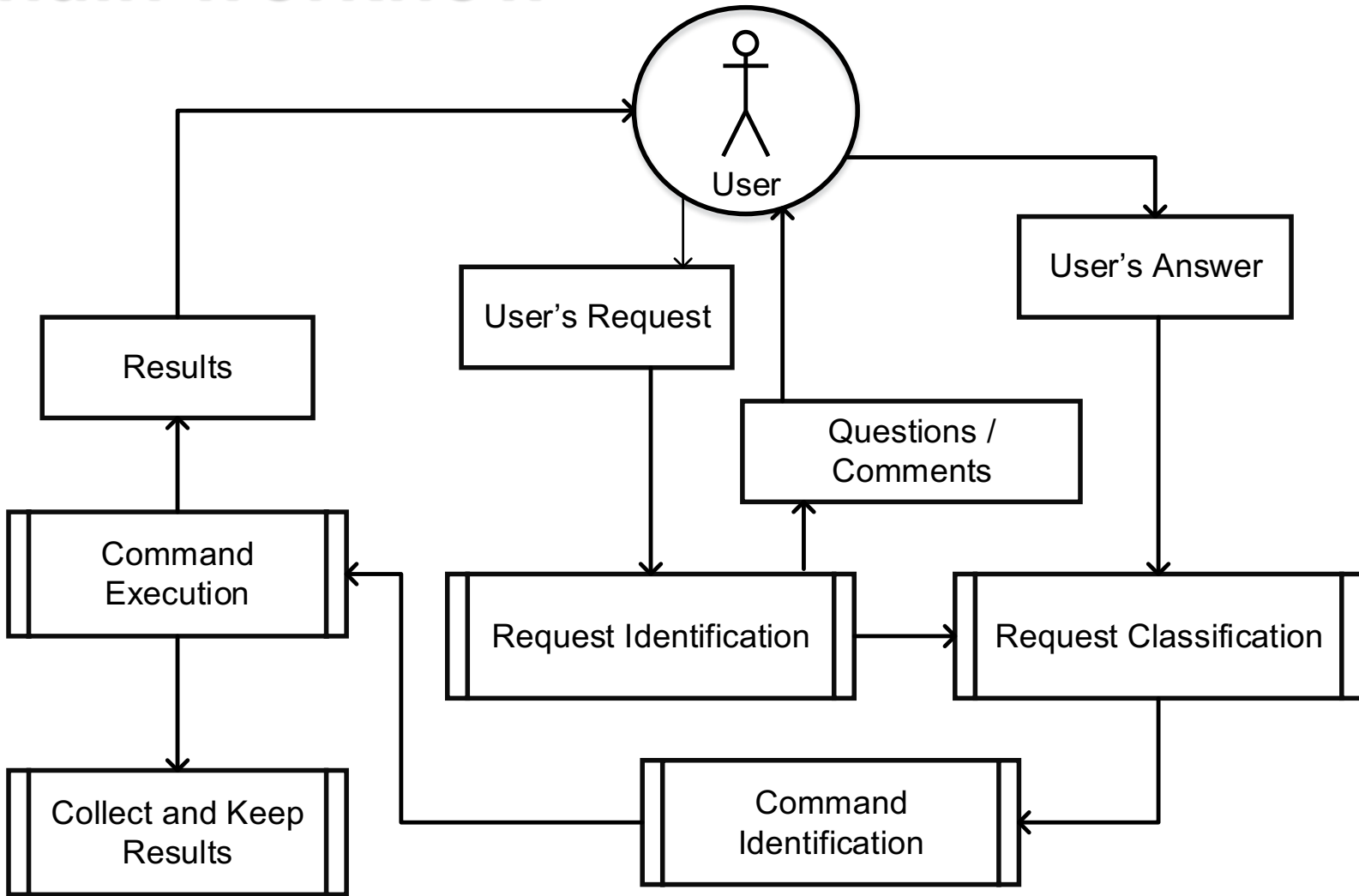
# Main components of CBI Virtual Assistant



- **Natural language conversation processing:** is responsible for processing the dialogue between the user and the virtual assistant.
- **Task identification:** involves identifying the commands that best match the user's needs based on natural language processing and semantic analysis of the input text.
- **Recommender model:** is responsible for personalizing the data exploration experience based on the user's preferences, data analysis history, and behavior.

The goal of a virtual assistant is to make data exploration more accessible to a wider range of users and to reduce the time and effort required for data analysis

# The main workflow



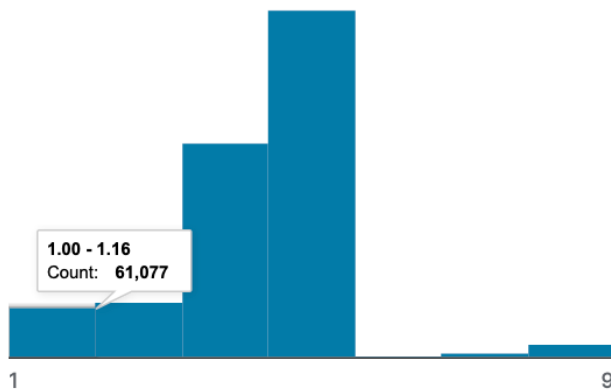
# Dataset for experimenting

- ▶ Dataset of road accidents in France (**Bases de données annuelles des accidents corporels de la circulation routière - Années de 2005 à 2021**)

<https://www.data.gouv.fr/en/datasets/bases-de-donnees-annuelles-des-accidents-corporels-de-la-circulation-routiere-annees-de-2005-a-2021/#>

- ▶ Example: Road category (1 – Motorway, 2 – National road, 3 – Departmental road, 4 – Communal roads, 5 – Outside the public network, 6 – Car park open to public traffic, 7 – Urban metropolis roads, 9 – other way )

# catr



Valid	840k	100%
Mismatched	1	0%
Missing	0	0%
Mean	3.42	
Std. Deviation	1.21	
Quantiles		
	1	Min
	3	25%
	4	50%
	4	75%
	9	Max

# Exploring Accidents in France

## ✓ Exploration based on date of accidents

- Is the number of accidents per year decreasing ?
- Which Day-of-the-Month is most safe to drive ?
- What weather has the most accidents?

## ✓ Exploration based on roads where accidents occurred

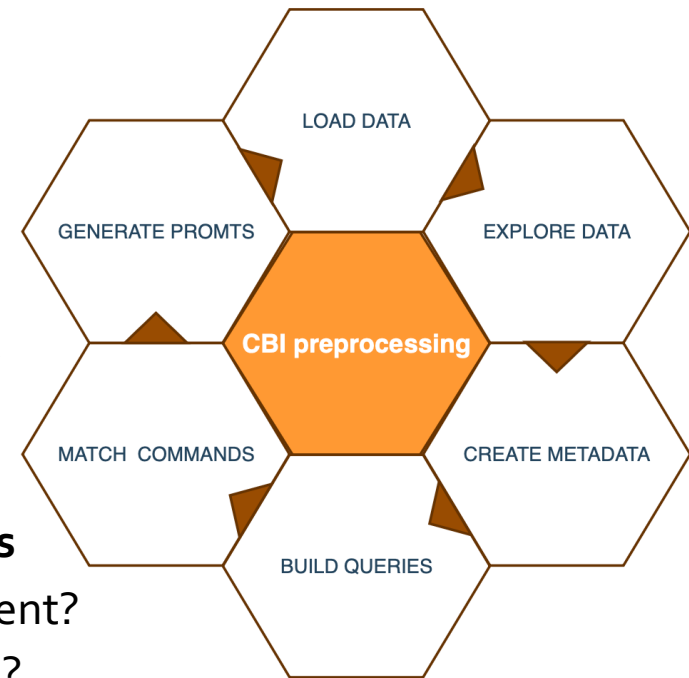
- Which types of roads are high risk?
- Which type of road gradient is high risk?

## ✓ Exploration based on people involved in the accidents

- What was the condition of the people after the accident?
- What was the age distribution of the people involved?
- What was the sex distribution of the people involved?

## ✓ Exploration based on use of safety equipment

- What was the distribution of Safety Equipment used?
- Did use of Safety Equipment impact condition of people after the accident?



# What weather has the most accidents?

## Conversation (example):

What weather has the most accidents?

Do you want to explore an atmospheric conditions related to the accidents?

Yes, I do

For what period of time?

For all available data

Okay

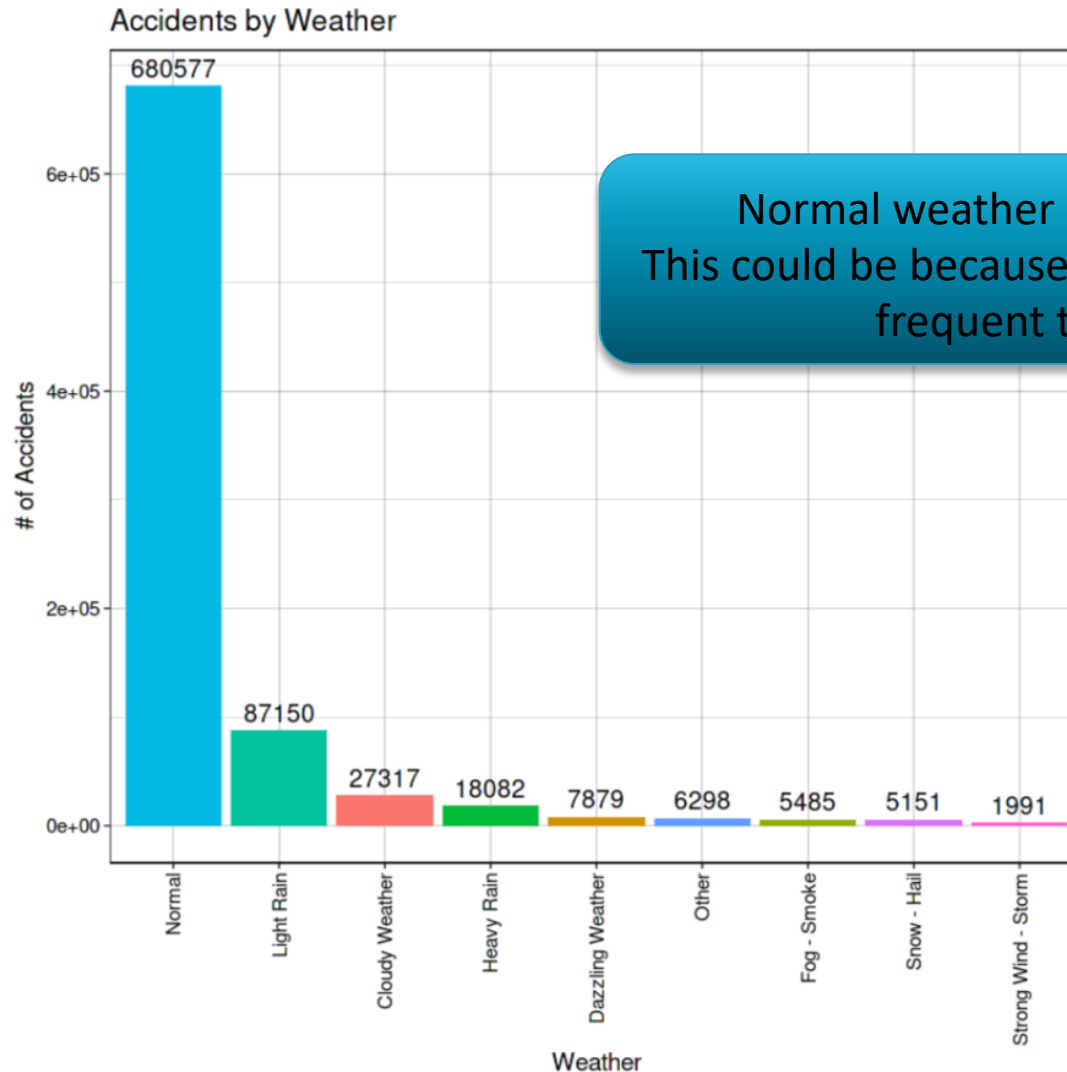
The most accidents were under Normal atmospheric conditions. There are 680571 accidents or 81 %

## Metadata (example):

- ▶ Column\_name = 'atm'
- ▶ Column\_description = 'Atmospheric conditions'
- ▶ Column\_value\_defined =  
[1 - Normal,  
2 - Light rain,  
3 - Heavy rain,  
4 - Snow-hail,  
5 - Fog - smoke,  
6 - Strong wind-storm,  
7 - Dazzling weather,  
8 - Cloudy weather,  
9 - Other ]

# What weather has the most accidents?

```
clear_data['atm'].value_counts().idxmax()
```

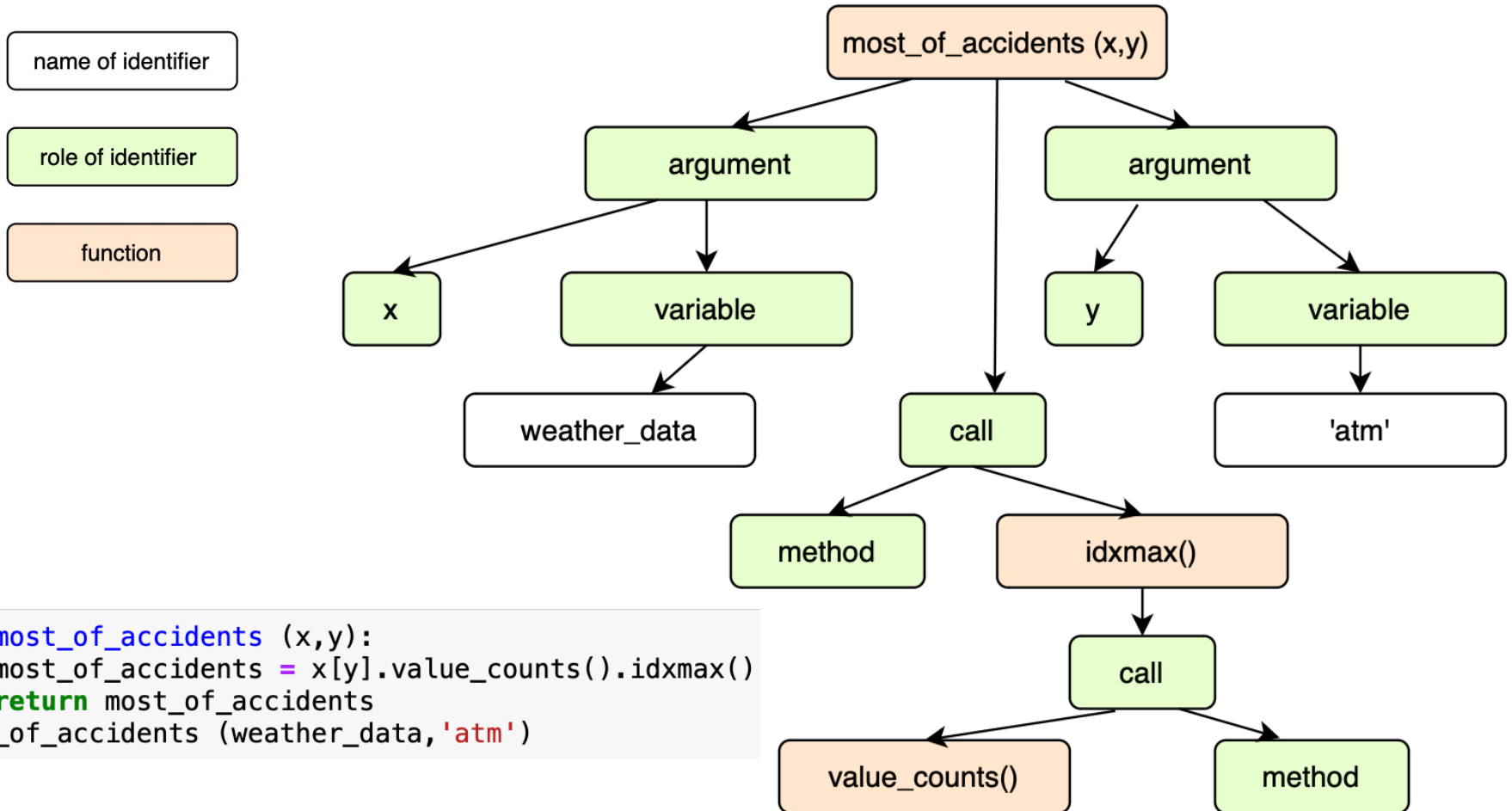


Normal weather has the most accidents. This could be because normal weather is the most frequent type of weather.

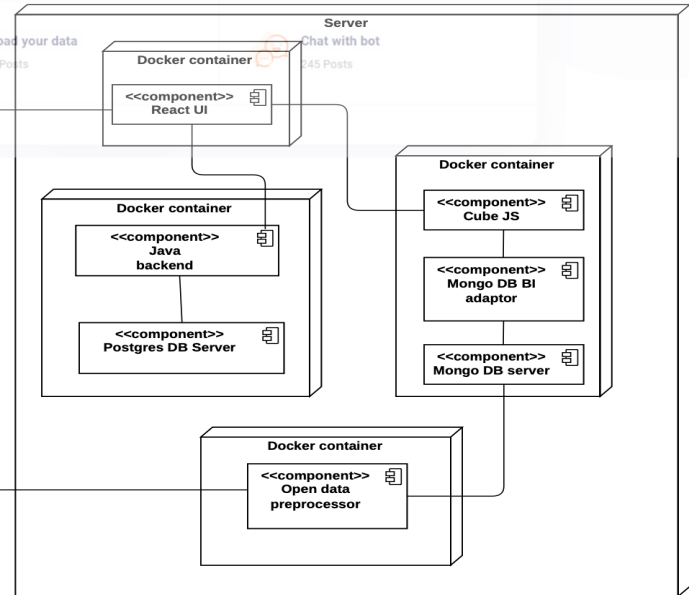
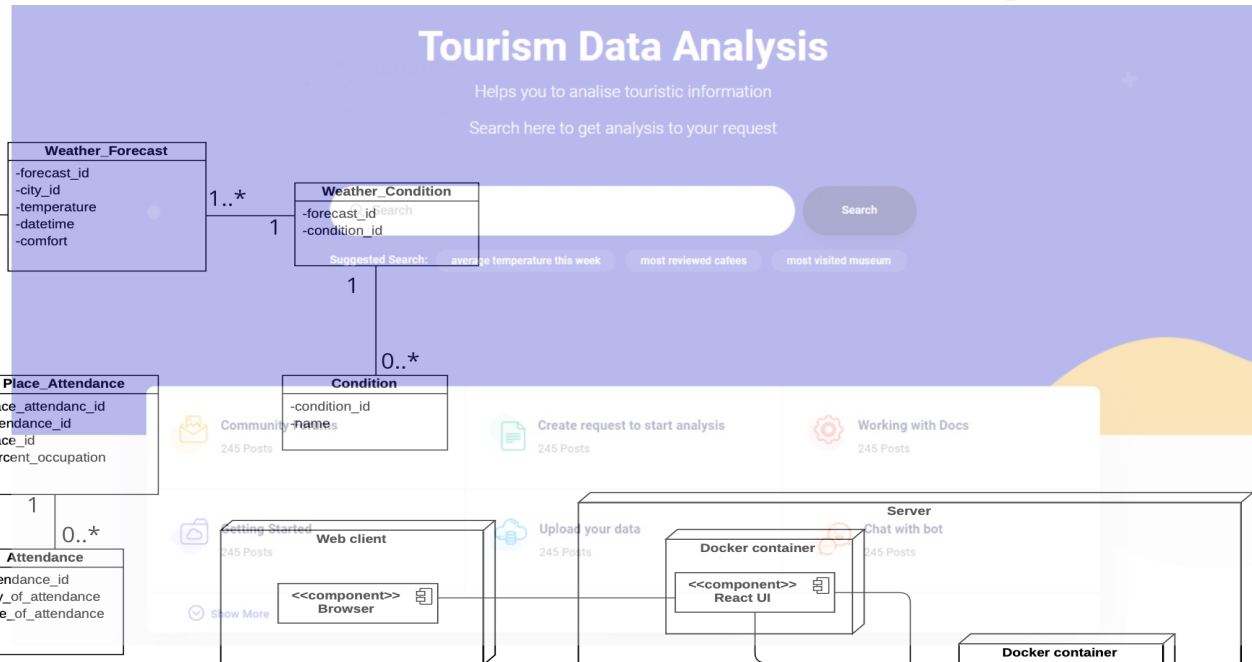
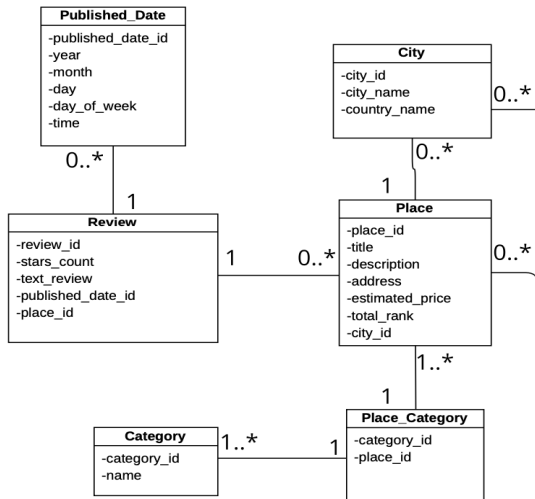
	weather	weather_n
1	Normal	680577
2	Light Rain	87150
3	Heavy Rain	18082
4	Snow - Hail	5151
5	Fog - Smoke	5485
6	Strong Wind - Storm	1991
7	Dazzling Weather	7879
8	Cloudy Weather	27317
9	Other	6298



# Abstract syntax tree creating



# Tourism Domain as a case-study

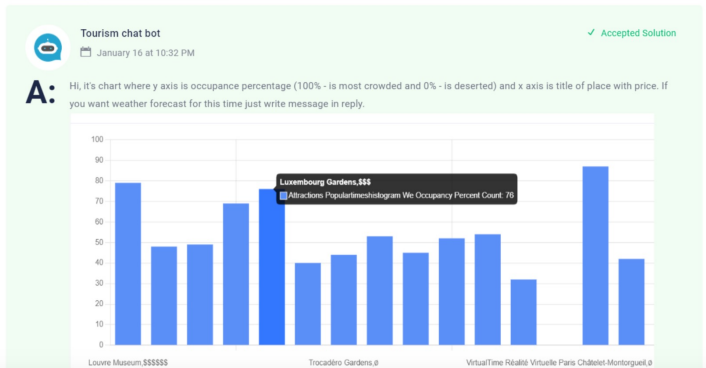


**R:** How much is crowded Luxembourg Gardens in Paris and its price comparing to another popular Paris attractions

I want to see how crowded Luxembourg Gardens usually on Wednesday and average price of entrance comparing to another popular Paris attractions

Paris, Tourism, Attractions

Reply I have this question too (20)



# Tourism Domain as a case-study

BI4People Tourism app

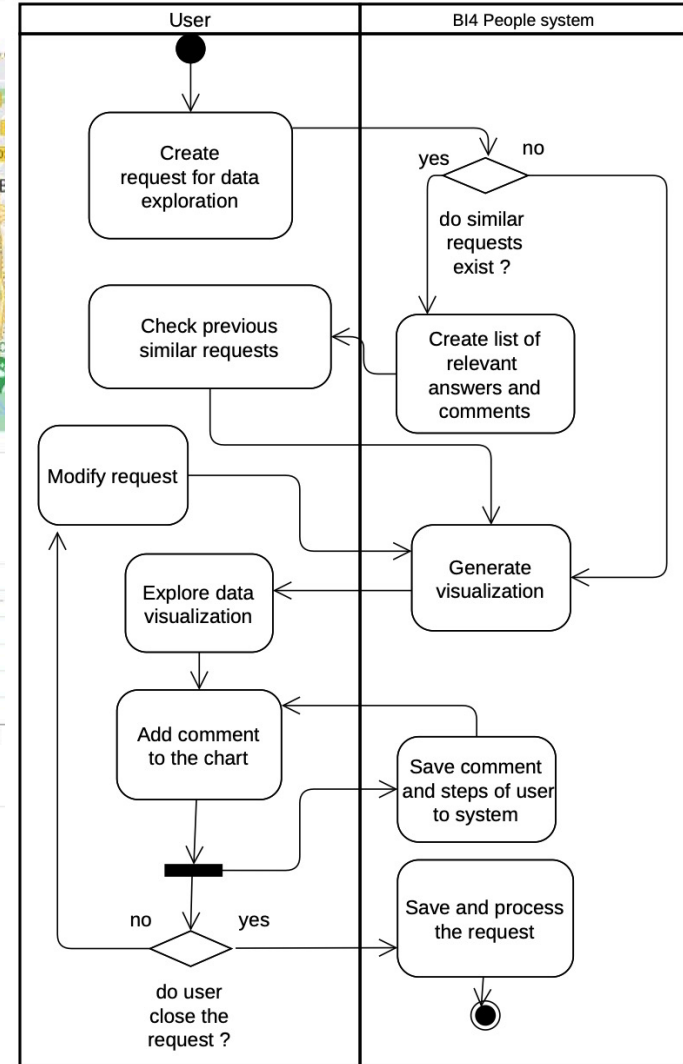
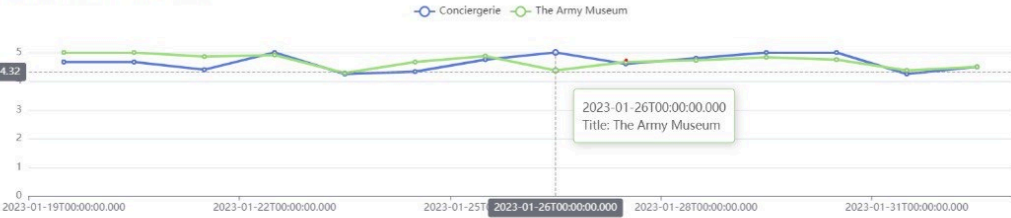
Best places to visit near the city center



Average occupancy by category and day of week



Review distribution of places in January



# References

- ▶ **Towards Collaborative Business Intelligent Framework: Crowdsourcing Approach** // 42nd International Conference on Organizational Science Development, editors Polona Šprajc ... [et al.]. - 1st ed. - E-book (available at <https://press.um.si/index.php/ump/catalog/book/768>). - Maribor : University of Maribor, University Press, 2023 -p.197-207
- ▶ **Reference Model for Collaborative Business Intelligence Virtual Assistant.** CEUR Workshop Proceedings, Vol-3403. 2023. p.114-125. url: <https://ceur-ws.org/Vol-3403/paper9.pdf>
- ▶ **Towards Cross-Lingual Transfer Based on Self-Learning Conversational Agent Model.** CEUR Workshop Proceedings, Vol-3396. 2023. p.194-205. url: <https://ceur-ws.org/Vol-3396/paper16.pdf>
- ▶ **Business Intelligence Virtual Assistant**– *accepted* in MoMLeT+DS 2023 *proceedings*
- ▶ **Recommender Chatbot as a Tool for Collaborative Business Intelligence in Tourism Domain** – *accepted in the PeRS@ADBIS'23 proceedings*